

Abstract of the Disclosure

Disclosed is a method and device for storing information about Web documents such as pages or sites in a manner which may be used in conjunction with inverted term lists to facilitate the retrieval of documents of interest from the Web. The method involves constructing compressed surrogates for documents, such that various operations may be performed without the need to retrieve a copy of the document from the Web. The method permits the efficient updating of inverted term lists when documents on the Web have been modified or deleted, and also permits the efficient processing of search queries in a variety of circumstances.

367238.1